Spectral Techniques for the Analysis of Tidal Time Series

By M. Bozzi Zadro and G. Poretti

Summary – Three different methods are given for the Fourier analysis of time series leading to the determination of high reliable values of frequencies, amplitudes and phases of the inherent components. For two methods the hypothesis is made that the time series is sufficiently long to separate very close components, while for the third the frequencies are supposed to be given. Two of these methods have been applied to a time series obtained by sampling an artificial tidal function.

1. – The spectral analysis methods commonly used during the last decade with the powerful help of large electronic computers, have been sometimes criticised. The reliability of the results seems to be strongly conditioned from the periodicity of the phenomenon and consequently from the possibility to develop it into harmonical components whose frequencies are integer multiples of a fundamental one. It follows that most of the natural almost-periodic phenomena as for example the tidal ones could not be analyzed by means of the spectral techniques.

The objections are justified if the spectral analysis methods are applied without a previous study on their limits of reliability. Such limits however, mainly depending on the data of the problem, can be often suitably widened as will be discussed in the following paragraphs.

For tidal data the classical harmonical analysis methods are usually preferred to Fourier ones. Harmonical analysis as known consists of the application of particular high selective band-pass filters each one extracting one single tide component. The small energy contribution due to the other tide components is neglected or taken into account by particular techniques.

The classical methods of harmonical analysis seem however to have some faults (Jobert [5]).

It can be noticed that:

1) They do not give the global vision of the phenomenon; energies eventually existing on frequencies different from the theoretic ones cannot be discovered when working on sea tides data, as information is generally lost both on interaction phenomena between tide components (shallow water tides), and on the seiches frequencies.

1) Istituto di Geodesia e Geofisica, Università di Trieste, 7, Via dell’ Università, 34100 Trieste, Italy.
2) Numbers in brackets refer to References, page 26.
2) The applied filters are generally built under the basic condition that their transfer function be unitary or in any case known at the requested frequency (or at a set of close frequencies) and almost zero for the other most important tide frequencies. This means that the eventual contributions of the above mentioned components are ignored and can moreover alter the results at the resolved frequencies.

3) A large set of papers, published during the last decade and dealing with time series analysis, and in particular with filtering and spectral analysis techniques show clearly that according to analytical and statistical laws, the reliability of the results depends on the sampling interval, on the length of the available series and on the truncation of the weighting function if filtering processes are involved. The quantification in spectral terms of the consequent unavoidable errors has not always been clearly pointed out from the classical methods of harmonical analysis.

Going back to the spectral analysis methods, they give remarkable advantages for what concerns all three above points. Point 2) in particular raises no problem; moreover, the application of suitable ideal filters before transformation can always avoid contamination effects.

Concerning point 3) the errors introduced by the spectral analysis can be singled out and can be reduced to negligible quantities.

The next two paragraphs are devoted to an analysis of the spectral methods and of the errors in the results with particular regard to almost-periodic phenomena; the third one to the techniques that can be applied in order to reduce those errors for a chosen number of frequencies. In the last paragraph these methods are tested by analysing an artificial time series resembling a tide process.

2. Let \( M \) time series \( \{x^m(n \Delta t)\} \) be given, with

\[
\begin{align*}
  m & = 1, 2, \ldots, M \\
  n & = N(m) \ldots N N(m) \\
  \Delta t & = \text{sampling interval}
\end{align*}
\]

obtained by sampling \( M \) records of the same stationary ergodic phenomenon \( x(t) \), relative to different time intervals (also partially overlapped).

Let the recorded phenomenon be given from the superposition of \( I \) cosinusoidal components with (at present unknown) frequencies \( f_i \) \( (i = 1, 2, \ldots, I) \) and that because of the choice of the sampling interval every energy contribution beyond the Nyquist frequency can be excluded; noise effects can be likewise avoided by filtering the time series.

The problem is now to determine the complex spectrum \( X(i f) \) with the best possible precision making use of the estimates of the spectral values computed for the \( M \) time series.

If only the power spectrum is requested it can be computed through the mean autocorrelation function of the observed time series (BLACKMAN and TUKEY [1]).
It is well known that this method gathers in a statistically reliable way information about the analysed series, but in comparison with the direct transform methods gives a poor spectral resolution.

In the present paper the direct Fourier transform method will be considered. This gives for the \( m \)-th series an estimate \( X^m(i f) \) sampled between zero and Nyquist frequency with sampling interval

\[
\Delta f_m = \frac{1}{[N N(m) - N(m) + 1] \Delta t}.
\]

In order to restrict truncation effects in the spectral results, the time series are modulated by a suitable time window \( D^m(n \Delta t) \) symmetric to the central time. The time window usually applied is the one we will refer to later on as \( D^m_2(t) \) (von Hann function) sampled like the time series at intervals \( \Delta t \); for \( N(m) = -N N(m) \) it is given by

\[
D^m_2(n \Delta t) = \begin{cases} 
\frac{1}{2} \left( 1 + \cos \frac{\pi n}{N(m)} \right) & \text{for } n = -N N(m) \ldots + N N(m) \\
0 & \text{elsewhere.}
\end{cases}
\]

Its Fourier transform is the spectral window \( Q^m_2(f; \bar{N}(m)) \)

\[
Q^m_2(f; \bar{N}(m)) = Q^m_0(f) + \frac{1}{2} Q^m_0(f + \frac{1}{t_N}) + \frac{1}{2} Q^m_0(f - \frac{1}{t_N})
\]

where

\[
t_N = \bar{N}(m) \Delta t
\]

\( Q^m_0(f) \) is the spectral window of the box-car function

\[
Q^m_0(f) = \bar{N}(m) \frac{\sin \pi f t_N}{\pi f t_N}.
\]

Applying the Cooley Tukey algorithm for the fast Fourier transform (FFT) (COOLEY and TUKEY [2] and COOLEY, LEWIS and WELCH [3]) the phases are carried out with reference to the time origin \( t_0 = N(m) \Delta t \) corresponding to the first value of the \( m \)-th time series; assuming temporarily for sake of simplicity that time as origin, the modulating function is \( D^m_2(t) * \delta(t - t_N) \) sampled at intervals \( \Delta t \), being now:

\[
t_N = \bar{N}(m) \Delta t
\]

and

\[
\bar{N}(m) = \frac{1}{4}[N N(m) - N(m)].
\]

The function \( X^m(i f) \) we get from the spectral values sampled at intervals \( \Delta f_m \), applying the convolution theorem is given by

\[
X^m(i f) = [A^m(f) + i B^m(f)] * Q^m_2(f; \bar{N}(m)) e^{2\pi i f t_N},
\]

\]
where
\[ A^m(f) + i B^m(f) \] is the Dirac comb, complex spectrum of the series under investigation, without truncation effects and with phases referred to the time \( t_0 \).
\[ Q_2^m(f; \bar{N}(m)) e^{2\pi i f t_N} \] is the complex spectrum of the time window shifted by \( t_N \).

In order to obtain the phases referred to the time \( t=0 \) instead of \( t_0 \), it has to be noticed that \( \frac{1}{2} \) factor a part:
\[ A^m(f) + i B^m(f) = \sum_{i=1}^{I} a_i e^{i\phi f f_i} \delta(f - f_i) \]
where
- \( a_i \) is the amplitude of the wave with frequency \( f_i \)
- \( \phi f f_i \) is the corresponding phase referred to \( t_0 \) time.

The complex spectrum referred to the time \( t=0 \) will be given from
\[ A(f) + i B(f) = \sum_{i=1}^{I} a_i e^{i(\phi f f_i - 2\pi f f_0 t_0)} \delta(f - f_i) \] (1)
where
\[ \phi f f_i = \phi f f_i - 2\pi f f_i t_0 \]
is the phase we are looking for.

Referring to the spectral values \( X^m(i f) \), they can be interpreted as follows
\[ X^m(i f) = \left[ A(f) + i B(f) \right] \ast Q_2^m(f; \bar{N}(m)) e^{2\pi i (f f_0 + f t_N)} \]
\[ = \sum_{i=1}^{I} a_i e^{i\phi f f_i} Q_2^m(f - f_i; \bar{N}(m)) e^{2\pi i [f f_0 + (f-f_i) t_N]} \] (2)

3. A first possible misinterpretation of the single spectrum \( X^m(i f) \) lies in the overlap of adjacent spectral windows centered on close frequencies. Let \( f_i \) and \( f_{i+1} \) be two close frequencies, being \( \delta f = f_{i+1} - f_i \). Neglecting the contributions of the side lobes of the spectral window (i.e. considering as vanishing the function \( Q_2^m(f; \bar{N}(m)) \) when \( |f| \geq 2 A f_m \)) it follows that the energies associated with the \( f_i \) and \( f_{i+1} \) frequencies will be correctly recognized only when \( \delta f \geq 4 A f_m \).

This means that if the size of the whole time interval can be properly chosen, energies belonging to close frequencies can always be separated; if otherwise the recording is limited (or with many interruptions) but if the frequencies are known the contamination effects must be taken into account.

Excluding the contamination between neighbouring spectral windows the problem is now the identification of frequencies, amplitudes and phases, as well as the determination of the inherent errors.

The frequencies \( f_i \) are normally identified as corresponding to the maxima of the modulus \( |X^m(i f)| \) at the resolved frequencies: therefore the largest error on the frequencies is \( \pm \frac{1}{2} A f_m \) i.e. inversely proportional to the length of the time series.

Taking into consideration the values of the spectral window between \( \frac{1}{2} A f_m \) and \( \frac{3}{2} A f_m \) the largest relative error on the amplitude is up to 15%, the computed value
being always less than the real one. Notwithstanding this, assuming that within the interval \(-\frac{1}{3} \Delta f_m + \frac{1}{3} \Delta f_m\) the unknown frequency \(f_i\) follows the uniform probability distribution law and taking into account the values of the spectral window, the relative error on the amplitude is up to 10% with an 80% probability.

The error is much more remarkable on the phases; in fact, if the frequency \(f_i\) lies between the \(f_k\) and \(f_{k+1}\) resolved frequencies, from formula (1) we get the following

\[X^m(i, f_k) = a_i e^{i(\varphi_i + 2\pi(f_{k0} + (f_i - f_k) t_0))} Q^\pi(f_k - f_i; N(m))\]

for \(f = f_k\) and a similar one for \(f = f_{k+1}\).

The phase computed as arctangent of the ratio between the sine and cosine Fourier transform at \(f_k\) frequency gives instead of the requested value \(\varphi_i + 2 \pi f_i t_0\) (from which we would get \(\varphi_i, f_i\) and \(t_0\) being known),

\[\varphi_i + 2 \pi f_i t_0 + (f_k - f_i) t_n\]

Consequently the estimated phase varies linearly with the frequency between \(f_k\) and \(f_{k+1}\): it can be easily seen that the phase difference between the two resolved frequencies is \(\pi\). Taking the phase value corresponding to the resolved frequency nearest to \(f_i\) the error can be up to \(\pi/2\); like the above mentioned error on the amplitude the phase error is independent from the length of the time series.

In the analysis of almost-periodic phenomena as the tidal ones, the \(I\) frequencies \(f_i\) cannot match so many resolved frequencies, no matter how we change the length of the series. It follows that especially the phase errors cannot be checked by the application of the simple Fourier analysis without taking into account the transform process.

Three different methods will be examined on the base of the foregoing theory, which allow highly reliable estimates of the spectrum \(X(i, f)\) in correspondence of a set of \(I\) frequencies either known or unknown.

4. – Among the three methods that will be examined, the first and the second one assume known only with a rough approximation the \(f_i\) frequencies and need sufficiently long time series in order to avoid contamination phenomena due to neighbouring spectral windows. The third method assumes as exactly known the frequencies under investigation and can be used with series of any length.

The first method will be mentioned very shortly, having been rejected for the applications because of the large computational work and the consequent computer time required.

According to this method a set of \(M\) complex spectra \(X^m(i, f)\) is carried out for each frequency \(f_i\) only roughly known (or identified through the spectral analysis); the set is relative to \(M\) time series obtained increasing each time the number of data so that the first and the last one differ of about one complete period of the wave under investigation.
This way, in the neighbourhood of $f_i$, in an interval corresponding to the elementary band $Af_m$ we get $M$ equally spaced estimates of the spectral window centered at frequency $f_i$ itself.

This method, if properly used, gives a remarkable precision, i.e.: frequency and phase are determined with errors $M$ times smaller than those obtained from a single analysis whereas the amplitude is determined with a relative error of less than some units per cent when $M$ is greater than 5.

Only one set of $M$ spectra (for example 24, if hourly sampling is given) is sufficient in the case of tides, for the analysis of diurnal, semidiurnal and terdiurnal components, the higher frequencies being approximately integer multiples of the diurnal ones. Anyway, for the previously expressed reasons the following methods have been preferred.

The second method makes use of a single very long time series (therefore $M=1$) in order to get an elementary frequency band as narrow as necessary to avoid the superposition of neighbouring spectral windows.

Being $M=1$, let us set for simplicity $N(1)=0$ and $N N(1)=N N$ and let us consider in detail the $X(i f)$ spectral values at the resolved frequencies $f_k$ and $f_{k+1}$; let be $f_k < f_i < f_{k+1}$ where $f_i$ is the unknown frequency whose associated amplitude $a_i$ and phase $\phi_i$ are also unknown. Let the energy within the frequency band $f_k - f_{k+1}$ have been discovered through an inspection on the amplitude spectrum; otherwise it could be expected from the knowledge of the phenomenon.

From equation (1) with obvious simplifications of symbols we get

$$R(f_k) = a_i Q_2(f_k - f_i) \cos[\phi_i + \pi(f_k - f_i) N N \Delta t]$$
$$I(f_k) = a_i Q_2(f_k - f_i) \sin[\phi_i + \pi(f_k - f_i) N N \Delta t]$$

where

$$X(i f_k) = R(f_k) + i I(f_k);$$

similar formulas can be written for $f_{k+1}$.

This way two estimates $a'_i$ and $a''_i$ of the amplitude $a_i$ can be obtained at $f_k$ and $f_{k+1}$

$$a'_i = \sqrt{R^2(f_k) + I^2(f_k)} = a_i Q_2(f_k - f_i)$$
$$a''_i = \sqrt{R^2(f_{k+1}) + I^2(f_{k+1})} = a_i Q_2(f_{k+1} - f_i).$$

Similarly two estimates $\phi'_i$ and $\phi''_i$ of the phase $\phi_i$ (notice that $\phi''_i - \phi'_i = \pi$) can be obtained:

$$\phi'_i = \arctg \frac{I(f_k)}{R(f_k)} = \phi_i + \pi(f_k - f_i) N N \Delta t$$
$$\phi''_i = \arctg \frac{I(f_{k+1})}{R(f_{k+1})} = \phi_i + \pi(f_{k+1} - f_i) N N \Delta t.$$  

Formulas (3') and (3'') give a system of two equations with two unknowns $a_i$ and $f_i$; it can be solved with a method of iterated approximations. Knowing $a_i$ and $f_i$ the
phase $\varphi_i$ can be calculated by $(4')$ or $(4'')$. It must be understood that as far as the phases are concerned, more reliable results can be reached through formulas $(4')$ and $(4'')$ where the $f_i$ frequencies are exactly known.

The third method makes use of $M$ complex spectra obtained from $M$ time series of whatsoever length (neighbouring spectral windows can be partially overlapped); the $I$ frequencies we want to examine in detail are given. This method is particularly suitable to the analysis of registrations that for instrumental reasons are subject to interruptions and whose frequencies are known in advance (as in the case of sea and earth tides).

The $I$ given frequencies are gathered in such a way that each group includes adjacent frequencies whose spectral windows are partially overlapped in the spectrum of largest elementary frequency band.

Let for example $x_i(t_n \Delta t)$ be the shortest time series of the set $\{x^m(n \Delta t)\}$ and $f_1 < f_{i+1} < \cdots < f_{i+s}$ be a group of $s+1$ frequencies, the difference between two successive of them being lower than $4 \Delta f_i$.

For the $m$-th spectrum $X^m(i \Delta f)$ the frequency bands $k_m \Delta f_m$, $(k_m+1) \Delta f_m$, ..., $(k_m+r_m) \Delta f_m$ are considered, where $k_m$ and $r_m$ are chosen so that

$$(k_m + 3) \Delta f_m < f_i < (k_m + 4) \Delta f_m$$

and

$$(k_m + r_m - 4) \Delta f_m < f_{i+s} < (k_m + r_m - 3) \Delta f_m.$$  

This way we are sure to examine all the frequency bands containing the $s+1$ frequencies and only those where the energy contributions are meaningful.

According to equation (2), a system of $2(r_m+1)$ equations and $2(s+1)$ unknowns can be written (HATZFELD [4]); in fact for the general $f_k$ resolved frequency, we get the two equations:

$$R^m(f_k) = \sum_{i=1}^{l+s} u_i Q^m_2(f_k - f_i; \bar{N}(m)) \cos 2 \pi (f_i t_0 + (f_k - f_i) t_n)$$

$$- \sum_{i=1}^{l+s} v_i Q^m_2(f_k - f_i; \bar{N}(m)) \sin 2 \pi (f_i t_0 + (f_k - f_i) t_n);$$

$$I^m(f_k) = \sum_{i=1}^{l+s} u_i Q^m_2(f_k - f_i; \bar{N}(m)) \sin 2 \pi (f_i t_0 + (f_k - f_i) t_n)$$

$$+ \sum_{i=1}^{l+s} v_i Q^m_2(f_k - f_i; \bar{N}(m)) \cos 2 \pi (f_i t_0 + (f_k - f_i) t_n);$$

where

$$R^m(f_k) + i I^m(f_k) = X^m(i \Delta f_k)$$

and

$$u_i = a_i \cos \varphi_i$$

$$v_i = a_i \sin \varphi_i.$$  

Thereafter $a_i$ and $\varphi_i$ can be carried out by knowing $u_i$ and $v_i$. 

A system of $\sum_{n=1}^{N} 2(r_m+1)$ equations with $2(s+1)$ unknowns is obtained if all the equations brought about from the $M$ complex spectra are taken into consideration. The problem can be solved like a problem of indirect observations by means of the least squares method.

The method allows a remarkable precision both of amplitudes and phases and an exact evaluation of the errors, including those brought about from the rough separation of close frequencies: for this purpose the computation of the correlation matrix is suggested.

5. – In order to have an *a priori* evaluation of the errors introduced in the case of tide analysis, both methods have been applied to the spectral values of a time series built with 7 among the largest diurnal and semidiurnal tide components.

According to the Cooley-Tukey algorithm for the fast Fourier transform, a power of 2 numbers of data is most suitable for computational reasons.

The storage capability of the available machine (IBM 7044 of the Computer Centre of the University of Trieste) allows us to use a maximum set of 8192 data.

In the first analysis these have been obtained by sampling the above mentioned artificial tide function every 21 hours. The arguments of this time series appear in Table 1 as input values.

Because of the 21 hours sampling interval, the diurnal energies will appear in the spectrum as a first folding, while the semidiurnal will result as a double folding around the Nyquist frequency (1/42 cph).

In order to avoid aliasing phenomena between diurnal and semidiurnal components the computations have been carried out in two different steps. The first time extracting the diurnal components, the second time extracting the semidiurnal ones by applying two band-pass ideal filters. The results are summarized in Table 1.

<table>
<thead>
<tr>
<th>Wave</th>
<th>Frequencies</th>
<th>Amplitudes</th>
<th>Phases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Input</td>
<td>Computed</td>
<td>Input</td>
</tr>
<tr>
<td>$O_1$</td>
<td>13.943036</td>
<td>13.943036</td>
<td>6.000</td>
</tr>
<tr>
<td>$P_1$</td>
<td>14.958931</td>
<td>14.958932</td>
<td>7.000</td>
</tr>
<tr>
<td>$K_1$</td>
<td>15.041069</td>
<td>15.041070</td>
<td>18.000</td>
</tr>
<tr>
<td>$N_2$</td>
<td>28.439729</td>
<td>28.439737</td>
<td>4.000</td>
</tr>
<tr>
<td>$M_2$</td>
<td>28.984104</td>
<td>28.984104</td>
<td>27.000</td>
</tr>
<tr>
<td>$S_2$</td>
<td>30.000000</td>
<td>30.000000</td>
<td>16.000</td>
</tr>
<tr>
<td>$K_2$</td>
<td>30.082137</td>
<td>30.082138</td>
<td>5.000</td>
</tr>
</tbody>
</table>

As far as the frequencies are concerned an error appears sometimes on the last significant figure which includes also the rounding error of the computer. The range of the error on the amplitudes is from 0.1% to 4% and is larger for smaller amplitudes. The error on the phases lies between .5 and .02 degrees.
Table 2

<table>
<thead>
<tr>
<th>Wave</th>
<th>Amplitudes</th>
<th>Phases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Input</td>
<td>Computed (A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.002</td>
</tr>
<tr>
<td>$O_1$</td>
<td>6.000</td>
<td>6.002</td>
</tr>
<tr>
<td>$P_1$</td>
<td>7.000</td>
<td>7.003</td>
</tr>
<tr>
<td>$K_1$</td>
<td>18.000</td>
<td>18.009</td>
</tr>
<tr>
<td>$N_2$</td>
<td>4.000</td>
<td>4.002</td>
</tr>
<tr>
<td>$M_2$</td>
<td>27.000</td>
<td>27.003</td>
</tr>
<tr>
<td>$S_2$</td>
<td>16.000</td>
<td>15.998</td>
</tr>
<tr>
<td>$K_2$</td>
<td>5.000</td>
<td>5.002</td>
</tr>
<tr>
<td></td>
<td>Input</td>
<td>Computed (A)</td>
</tr>
<tr>
<td></td>
<td>30°.000</td>
<td>30°.011</td>
</tr>
<tr>
<td></td>
<td>45.000</td>
<td>44.987</td>
</tr>
<tr>
<td></td>
<td>60.000</td>
<td>59.987</td>
</tr>
<tr>
<td></td>
<td>250.000</td>
<td>250.004</td>
</tr>
<tr>
<td></td>
<td>260.000</td>
<td>260.006</td>
</tr>
<tr>
<td></td>
<td>270.000</td>
<td>270.014</td>
</tr>
<tr>
<td></td>
<td>280.000</td>
<td>279.973</td>
</tr>
</tbody>
</table>

In the second analysis the time series has been sampled at 2 hours interval and 10 successive sections of 4096 data each have been separately analyzed. The results appear in Table 2.

In the column (A) the amplitudes and phases are given as obtained by the application of the least squares method to the spectral values of one time section alone.

In column (B), for a comparison, the results appear as obtained from the spectral values belonging to all 10 time sections.

Due to the width of the elementary frequency band ($Af=1/8192$ cph), the $K_1$ and $P_1$ waves have been analyzed together as well as the $S_2$ and $K_2$.

In the present case, since the data are free from noise, the errors on the values of the (B) column are almost of the same order of magnitude as those on the (A) column: they are up to $1\%$ for the amplitudes and lower than 0.04 for the phases.

References


(Received 5th June 1971)